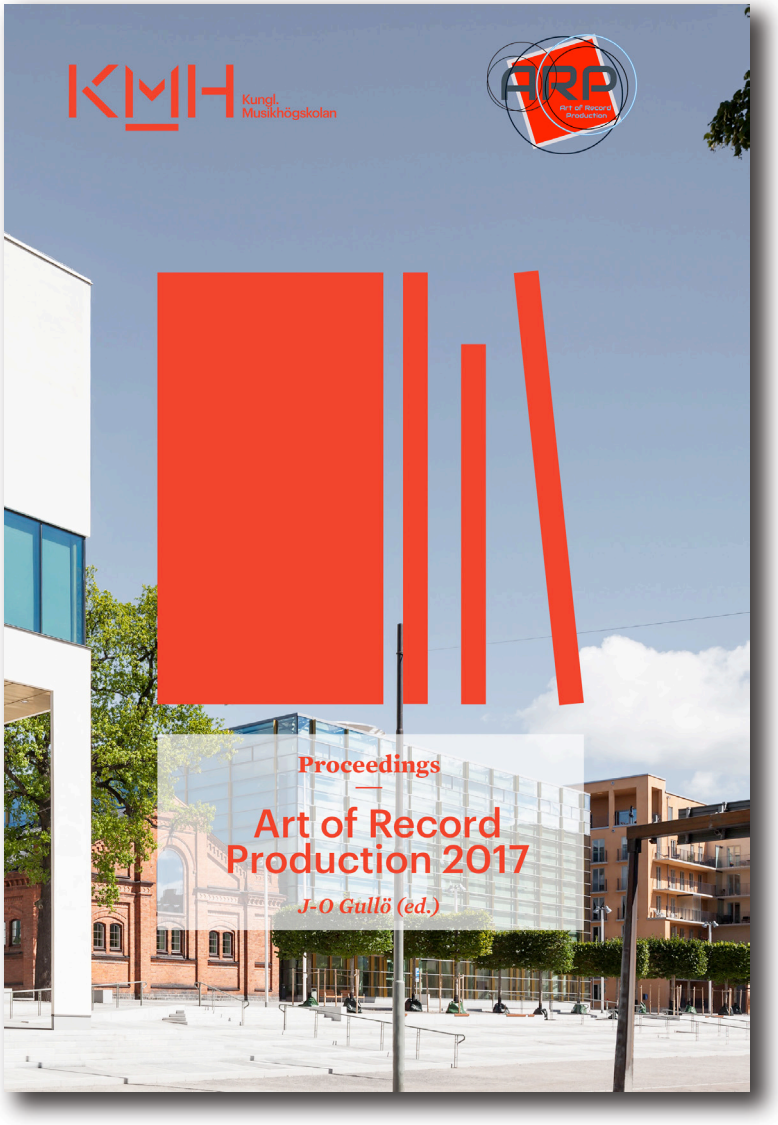
Proceedings of the 12th Art of Record Production Conference Mono: Stereo: Multi



Novotny, P. (2019). Creating an Immersive Fold-Out – Look Ahead. In J.-O. Gullö (Ed.), *Proceedings of the 12th Art of Record Production Conference Mono: Stereo: Multi* (pp. 205-224). Stockholm: Royal College of Music (KMH) & Art of Record Production.

ISBN 978-91-983869-9-8

Paul Novotny: Creating an Immersive Fold-Out – Look Ahead

Abstract

A comprehensive autoenthnographic case study, detailing the practice, engineering science and research used to create a stereo to 5.1-fold-out of the Canadian piano and bass jazz duo recording, *Look Ahead.* This paper asserts that a stereo to 5.1-fold-out, rather than fold-down is a preferable method because it affords each version a creatively distinct virtual acoustic environment by unique preparation. Fold-out results in two separate and unique masters that share a strong common foundation of ensemble sound. The conclusion reveals universal insights gleaned from the practice and research, henceforth assisting creators to adapt a stereo *sound-field* into a multichannel immersive *sound-scape*.

Introduction

Some years ago, I recall hearing my first playback of a jazz trio in surround-sound. I sat in the ideal listening position, looking ahead to the left, center and right speakers (L/C/R). The first thing I heard was a solo piano introduction, but surprisingly it came from the rear left and right speakers (LS/RS). As I turned around to face the music, the bass and drums entered from the L/C/R speakers which were now behind me. I was confused. My focus was not on the music because I was seeking a listening position that made the trio sound like a cohesive ensemble. In the end, I could not. That experience heightened my curiosity and I wondered what my music might sound like in 5.1, what process I would use to make it, and if others were asking these questions too.

This paper is an autoethnographic case study of the practice, engineering science and research involved in making the jazz duo recording, serendipitously titled, *Look Ahead*, recorded at 24 bit/96khz for playback in stereo and 5.1. It reveals the thinking, process and reasons for a stereo to 5.1-fold-

out, rather than a fold-down and it describes why fold-out is preferred by audio engineers for multichannel music production.[1]

I introduce terminology that differentiates stereo and surround-sound audio contexts with the hyphenated terms, *sound-field* and *sound-scape*. It seemed efficient to define the studio-made L/R stereo front image of a recording as a *sound-field,* while *sound-scape* refers to a studio-made 360-degree multichannel surround-sound audio image.

Sound-field versus sound-scape

Sound field (no-hyphen), created by L/R speakers in a room is a widely-used term but basically its job is to represent the nature of the sound we hear in that acoustic environment.[2] Components of the sound field include sound pressure level, anechoic or dry sound sources and their related perspectives of distance, horizontal position and room size. These traits are perceived from early reflections, reverb diffusion and spectrum texture in the recording but heard more simply as the performance staging and acoustic environment in recorded music. When music is constructed in a studio control room, the reflections of a virtual *sound-space* become very important because they indicate that the performance is occurring in either a natural environment or an exaggerated hyper-real environment. Alan Moore (Moore: 1992) described an abstract model called the soundbox – a 'virtual textural space' capable of replicating proximal and distal perception of depth spatiality as intimate, private, social and public.

The 2.0 sound field capably conveys these four perspectives through stereo L/R speakers, but like the human "field of view" and visual screen technology, the constructed L/R "sound field" is limited to the horizontal peripheral edges. Moore also uses the terms "vertical location" to express musical register (high and low notes) and "horizontal position" for L/R panning. William Moylan (Moylan: 2002, p.14-15) describes elevation thoroughly and says, "to date, the vertical plane has received little attention in audio because of playback format difficulties." However, current multichannel surround-sound for games, 360-degree sound for virtual reality (VR) and artificial reality (AR) are becoming more mainstream and so expansion of Moore's soundbox theory may now be appropriate. Use of the term "vertical

[1] Wolpert, Jeff, *Adjunct Professor of Music & Media,* wrote, *"in the engineering community, a stereo to 5.1-fold-out is an accepted practice for music production".*
University of Toronto, 11-19-17

[2] The British Society of Audiology provides information on international standards of sound fields used in audiometry. "ISO 8253-2 (1998) describes three types of sound field which are defined by the allowable variation of sound pressure level, produced by the output of a loudspeaker, in a small space surrounding a reference point. The reference point is roughly at the midpoint of the head of a hypothetical listener." It also goes on to define a free sound field as anechoic, a diffuse sound field where walls and ceiling have significant effect and a quasi-free sound field where the walls and ceiling have a moderate effect.

location" to express musical register has become unclear since elevation can now be well represented in multichannel playback formats such as Dolby Atmos, Auro 3D and DTS:X.[3]

The purpose of *Look Ahead* in 5.1 is to expand the limitations of the recorded virtual front L/R *sound-field* (hyphenated) into a virtual multichannel immersive *sound-scape,* thereby improving the perception of all "soundbox" traits. Like a TV monitor with more pixels, the SMPTE/ITU–5.1 (L, R, C, LFE, LS, RS) multichannel format provides improved positional definition, acoustic depth of field and spatial context because there are more speakers, thereby reducing the intrusion of diffused reflections caused by playback from speakers in an indoor space.

Since every *sound-source* is naturally mono, a 360-degree immersive *sound-scape* augments the expert listening experience with better spatial detail.[4] It can provide a cohesive virtual re-construction of an acoustic environment for studio-made music. A core difference between *sound-field* and *sound-scape* is that the *sound-scape's* virtual reflections and reverb can travel toward the listener and then pass behind him/her, hitting what would sound like a rear wall, acoustically locating the listener in the middle of a 360-degree virtual *sound-space*.

A challenge in construction of the 5.1 *sound-scape* is creating a believable transition from the L/C/R front to the LS/RS rear. Two helpful techniques are delay and a gentle lo-pass filtering of LS/RS signals. But when those techniques are folded-down (down-mixed from *sound-scape* to *sound-field*), unexpected problems in the perception of intimate versus public space can occur.

A static fold-down (down mix) collapses the LS/LR signals into the front L/R, creating unpredictable and undesirable spatial reflections, but a fold-out

[3] I checked both Moore's original 1992 book (located at the Toronto reference library) and the second edition available on Kindle (re-issued 2018), both refer to 'vertical height location' as 'musical register'. The most current technological development suggests this concept needs updating. At the SIRT, AES, SMPTE conference (Toronto, Pinewood Studios-02/21/18) senior manager of broadcast operations at Bell Media, CTV, Francis Nunan declared "The audio channel is dead, object audio is where we are headed." He stated that the reason for this is the need for a 'universal delivery payload' rather than the current system of 'versioned broadcast media.' This advancement will influence music playback as well as broadcast media because playback mediums (clubs, theatres, headphones etc.) are not standardized and rarely sound the same as control-room DAW mix playback. Gaming engines such as 'Unity' and 'Wwise' are already responsible for rendering spatialized audio in VR and AR, this now includes altitudinal height or elevation. Spatial rendering of audio is moving toward computer graphics processors, meaning that construction of Moore's "soundbox" traits and "textural strands" is becoming automated and informed by meta-data. As Moore states, "Technology and texture is an ongoing evolution. Technology limits what can be conceived." Object audio and immersive multi-speaker playback is evolving rapidly and expanding the limits, offering exciting possibilities for *sound-scape* creation and social reception.

[4] In his book, *Introduction to the Sociology of Music*, Theodore Adorno (1976: 3-14) describes an expert listener as "profoundly competent," "exceedingly rare" and "able to comprehend the multiple interrelationships present in music during the act of listening."

avoids these issues because both the 2.0 *sound-field* 5.1 *sound-scape* versions are uniquely prepared.

In the appendix, example one supports this observation by providing a side-by-side analysis test—5.1 fold-down to stereo compared to the prepared 2.0 mix—of the bass solo in *My Favorite Things,* illustrating this undesirable occurrence.[5] The LS/RS reflections were very appropriate in the *sound-scape* version, but not in the folded down *sound-field.*

The side-by-side fold-down of *My Favorite Things* also provided an opportunity to compare Loudness Units Full Scale (LUFS) and peaks. The result showed two masters that looked virtually identical according to the numbers, but sounded and felt very different. The fold-down test can be heard well using stereo headphones.

Research methodologies and report style

This paper utilizes a mixed research method. The recording was a heuristic arts-based improvisatory exercise that was guided by intuition and previous experience. The writing style of this case study is autoethnographic (Ellis, Adams, and Bochner: 2011, p. 273-290) and it attempts to organize the details of practice and research in the logical sequence of pre-production, production, post-production, mix and mastering. The primary questions were:

1. What, when, why and how did pre-existing multichannel formats originate?
2. What social reception was imagined for multichannel audio and what principles might have guided its creators to resolve their work into the eventual pre-existing playback contexts.
3. What universal insights support all immersive audio formats?
4. What principles guide the construction of acoustic environment and ensemble cohesion?
5. According to the history of multichannel sound, the first use of surround-sound as a storytelling device was pioneered by Walt Disney in the movie Fantasia. Multichannel sound employed psychological tension and release to heighten audience expectation and emotion using multichannel audio. This compelled me to look for a psychological and musicological explanation of 'tension and release' as it applies to the organization of sound.

[5] A-B Side-by-side analysis of 5.1-fold-down versus the actual stereo master. This example displays unpleasing reflections because of the fold-down. (See appendix, Example #1)

These questions suggested four specific research directions: 1) Multi-channel-sound history, 2) David Byrne's theory of "creation in reverse," (Byrne: 2012) 3) R. Murray Schafer's theories on acoustic environment, (Shafer: 1977) 4) David Huron's ITPRA theory, which describes the five stages of expectation, and reduces them to two experiences of *pre-outcome* and *post-outcome*, commonly known as tension and release. (Huron: 2006, pp. 33-34, iBooks) With reflection I gradually realized that development of Walt Disney's *Fantasound* was motivated by Disney's innate understanding of tension and release, or more specifically, David Huron's ITPRA theory. That notion led me to view all creators of *sound-field* and *sound-scape* as "choreographers of emotional expectation and experience", who instinctively employ ITPRA principles. My conclusion presents seven foundational insights that can guide decisions when creating a *sound-scape*. It also advocates for the use of a fold-out process to afford the project two uniquely prepared masters, that share and benefit from a common ensemble sound.

Research begins with a chronological overview of multichannel sound

The earliest example of two-channel sound (2.0) over wire comes from Clément Ader in 1878. He placed twelve telephone transmitters in left-right positioning at the stage of the Paris Opera, then ran wires through sewers for two miles to the Electrical Exhibition for reception of the sound. From 1881 to 1930 this stereo listening experience which Ader called "Binauriclar Audition*"* was commercialized as *Théatrephone* in France and *Electrophone* in the UK. Soon, in 1933 Alan Blumlein invented stereo binaural-sound and matrixing. Right from the beginning, two audio channels started engaging audiences and today, multichannel audio such as Dolby Atmos continues to heighten audience expectations and emotions in cinemas and night clubs, such as the *Ministry of Sound* in London. The pre-existing and future contexts of immersive audio continue to develop, and may soon be ubiquitous due to VR and AR. Table 1 provides a chronological overview showing when these and other key audio advancements came to be.

Table 1: The history of multichannel sound.

Date	Multichannel technological development
1878-81	First instance of two-channel sound over wire – Paris Opera, Clément Ader (Scientific American: 1881, Hertz: 1981 pp. 368-372)
1881-1930	*Théatrephone* and *Electrophone* are commercialized (Scientific American: 1881, Hertz: 1981 pp. 368-372)
1931-33	Alan Blumlein invents stereo binaural-sound and matrixing (Blumlein:1958) (Gerzon: 1992)
1934	Bell Labs experiments with 3 channel sound and "audio perspectives" (Pickering, Baender: 1953)

1940-41	*Fantasia*, Walt Disney introduces 6 track recording deployed in 5 channel *Fantasound,* (L, C, R, RS, LS) seeking an immersive audio audience experience. Multitrack/panning & overdubbing were invented in this era (Klapholz:1991) (Torick E. 1998)
1952-53	*Cinerama* (7 channel) and *Cinemascope* (4 channel) (Dientsfry: 2016)
1970-75	Sansui QS-Quadraphonic-Quintaphonic | Dolby matrix technology (Dientsfry: 2016) Ambisonics was developed. It is a full-sphere agnostic surround-sound technique[6]
1979	Dolby 5.1, pioneered by F.F. Copola and W. Murch - *Apocalypse Now* (Dientsfry: 2016)
1982	Dolby Surround (Julstrom: 1987)
1983	THX Ltd. A company founded by George Lucas that developed a high-quality assurance system and certified accurate reproduction through surround-sound playback systems
1985	Dolby AC-1[7], Matrixing (Julstrom: 1987)
1987	Society of Motion Picture & Television Engineers (SMPTE) accepts the 5,1 numeric-convention system (Dientsfry: 2016)
1981	AC-2
1991	AC-3
1992	Dolby SR*D
1993-94	93-Digital Theatre Sound (DTS) Sony Dynamic Digital Sound (SDDS) 94-Holophone Microphone system for immersive audio was invented[8]
1995	First AC-3 home theatre decoder
1998	First AC-3 5.1 sports broadcast
1999	Dolby Digital Surround EX-6.1 (*Starwars*) | Dolby E
2000	Dolby Pro-logic II
2004	Dolby Digital Plus (DD+)[9]
2012	Dolby Atmos (3D audio adding height). Approx. 300 theatre installations 2013.[10] SRS Labs "MDA" is a competing "object" based audio system to *Dolby Atmos*.[11]
2015-16	DTS:X, Object based immersive audio system that competes with Atmos. Announced in 2015, deployed in home AV receivers in 2016. 8K Super Hi-Vision currently testing by NHK-Japan, (Sigimoto, Takehero, Nkayama, Yasushige, Komori, Tomoyasu, Chinen, Toru, Hatanaka, Mitsuyuki: 2017)
2020	Predicted adoption of 8K Super Hi-Vision, (Sigimoto, Takehero, Nkayama, Yasushige, Komori, Tomoyasu, Chinen, Toru, Hatanaka, Mitsuyuki: 2017)

[6] Ambisonics: http://www.ambisonic.net/

[7] Dolby AC-1 was the first digital coding technology, AC-2 had better audio quality and AC-3 included 5.1 channels at a bitrate of 320 kbp/s. AAC is part of the MPEG 4 codec and utilises lossy compression. Lossy compression removes data to make the file size smaller and it compromises definition.

[8] In 1994, Canadian inventor Michael Godfrey achieved "the most realistic listening experience" with his Holophone microphone. It uses eight microphones placed in a dummy head and is regarded as the finest surround-sound microphone available.

[9] Dolby Pro Logic I decode technology started in 1987. It is frequency limited, matrixed 4.0 surround-sound derived from a stereo mix. The surround-channel is slightly delayed keeping attention focused forward. Pro logic II provides 5 full frequency channels, IIx provides 6.1 & 7.1 capability and IIz provides height information to support Dolby Atmos. Dolby Digital AC-3 supports 5 full-bandwidth channels at 640 Kbit/s but DD+ supports up to 15 full-bandwidth channels at 6.144 Mbit/s.

[10] Dolby Investor Relations, http://investor.dolby.com/releasedetail.cfm?ReleaseID=799045 10,22,13

[11] SRS Labs developed the *Sound Retrieval System* technology. In 2008, approximately 36 million *SRS* equipped flat-screen TV's were shipped. In 2012 *SRS* was acquired by *DTS* Inc.

Byrne's "pre-existing context"

Because the history of multichannel film sound identifies what technological audio formats exist and how they originated, thoughtful creators can now consider them to be pre-existing music playback contexts. David Byrne writes, "I had a slow-dawning insight about creation. That insight is that pre-existing context largely determines what is written, painted, sculpted, sung or performed." He calls the insight "creation in reverse." (Byrne: 2012, p.18) In my opinion, *sound-field* and *sound-scape* are separate, virtual, pre-existing contexts, capable of enlivening audience engagement with their unique sound. However, to be most effective, they each require specific creative preparation, not adaptation using a static fold-down process.

Image 1: Performance oriented set up, the bassist is close to the piano keyboard.

Murray Schafer: the sound we hear

When describing natural sound, Composer R. Murray Schafer contends that *"...*outdoor sounds are different than indoor sounds,*"* and he uses the term "soundscape" to discuss the acoustic environment as the sound we hear. (Schafer: 1977) Byrne also observes that an outdoor stage and indoor concert hall are opposite acoustical environments that can uniquely inform creative decisions. He asserts that the creative path musicians unconsciously take fits into these pre-given listening contexts. Such insights by Byrne and Schafer helped me to understand that the *sound-field* should be uniquely adapted into *sound-scape* in order to be heard as a convincing virtual acoustic environ-

ment. I will discuss David Huron's ITPRA theory in the upcoming section titled *Reflective Thoughts.*

The Practice: constructing Look Ahead, aesthetic essence and pre-production

Establishing the aesthetic essence of a recording is the foundation of its ensemble sound. The performance aesthetic we established was based on an intimate, improvisational playing style employed by Oscar Peterson, i.e., the bass was positioned close to the piano keyboard. In addition, shorter reflections in the headphone *sound-space* complimented the spatial aesthetic of our relaxed home studio stage.[12] (See Image 1)

What's more, the sonic aesthetic was enriched using a diverse combination of esoteric microphones, pre-amps and Universal Audio (U.A.) plugins (tab. 2).

Table 2: Esoteric microphones, pre-amps and plugins provided our sonic aesthetic.

Instrument	Microphone	Pre-amp	Convertor	Plug-in	Preset
Bass	AEA R88	AEA-TRP	Apollo 16	none	none
Bass	AKG-C-391	Manley Vox Box	Apollo 16	U.A. Studer A800 printed	GP9 30ips Noise off
Bass direct (D.I.)	Underwood bridge pickup	Summit Audio TD100 to a U.A. 6176	Apollo 16	U.A. Studer A800 Manley Massive Passive printed	GP9 30ips Noise off 2K notched out
Piano "in"	Calrec-SoundField	SPS 422B	Apollo 16	U.A. Studer A800 printed	GP9 30ips Noise off
Piano "out"	2 U87's	Langevin Dual Vocal Combo	Apollo 16	U.A. Studer A800 printed	GP9 30ips Noise off

Repertoire and performance

Pianist Robi Botos provided the lead sheets for *Praise* and *Budapest* while I contributed lead sheets for *A Gentle December Day* and *Porters Hymn*. These original songs shared common musical traits—they were slow in tempo, contemplative and quiet, with dynamics ranging from *ppp to mf*, features that we believed would enhance the way our *sound-source* touched the virtual *sound-space* in stereo and 5.1.

[12] Conversely, Keith Jarret situates the bassist at the far end of the piano.

Sound-field leads to Sound-scape-recording the bass and piano

The *Look Ahead sound-field* begins at the center position of bass, recorded with a carefully centered stereo ribbon microphone, a mono hyper-cardioid condenser and a direct bridge pickup signal (D.I.). The tall stereo ribbon mic was placed in an acoustic shield to reduce rear reflections and focus the intimate stereo L/R spatial *sound-field* of the bass (image 2).

Image 2: Microphones used to record the upright bass.

The piano mic technique utilized two outside Neumann U87 microphones, providing a focused center image that blended with an inside-placed stereo Calrec-SoundField (C-SF) microphone. We preferred the U87s for headphone monitoring because the C-SF was too intimate, but still necessary. *(See discussion of piano panning and processing is in the mix techniques section.)*

Reverbs

An artificial *sound-space* was created using multiple instances of stereo and 5.1 reverbs, mixed and panned to overcome dead-spots between the front right to right side (R/S) and front left to left side (L/S) channels. Since there were no drums, the *sound-space* became an active participant in the ensemble, responding with dynamic spectral excitement. The 5.1 mix used the stereo 'dry' ensemble as its *sound-source* to ignite the stereo and 5.1 reverb chambers, creating a responsive and immersive virtual *sound-scape* that could be savored in the center listening position.[13]

[13] Stereo playback in a room is always 3 dimensional due to room reflections, it is never planar. The listening location within a room can be referred to as a Euclidean space, defined by Cartesian coordinates.

Post production and mix template

We tracked into Pro Tools with Universal Audio (U.A.) processing. Studer A800 tape saturation provided warm saturation and this workflow decision afforded us more computer processor power in the mix stage. A film-style mix template that accommodated discrete stereo and 5.1 mixing simultaneously, was ideal for the fold-out. It provided easy sub-grouping and bussing of individual tracks to 5.1 sub-masters and the LFE.[14] It must be understood that a fold-out from a fixed stereo master-print, known as *faux 5.1*, is not ideal because individual instrument tracks cannot be accessed. The mix template provided instrument sub-masters that allowed us to use gentle compression at several stages, gradually increasing levels while preserving a natural and warm character to the mix. Efficient solo and mute functionality allowed for speedy edit auditioning and easy analysis of blend relationships when joining individual reverbs into one coherent *sound-space* (image 3).

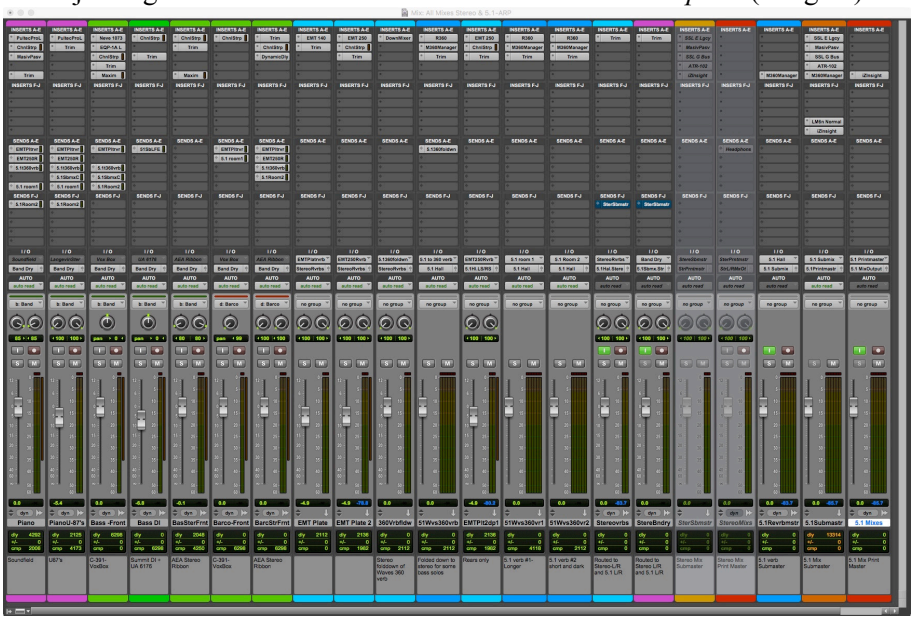


Image 3: Mix template showing routing and plug-ins.

Routing

The stereo piano and bass were routed to their own instrument sub-masters, which then led to a dry-band stereo sub-master. Three stereo reverbs (U.A. EMT-140/250, Waves R-360 5.1-folded-down to stereo) were bussed to their

[14] The Ability to compress, equalise, send or effect tracks or groups and any stage of routing is ideal. Only the bass D.I. (direct-signal) was routed to the LFE to make the fundamental deeper. Low frequency roll-off was applied at the sub-master stage using the Waves 360 sound manger. Roll off began at 80 HZ at 24 db per octave finishing at about 240HZ. The LS, RS were rolled off from at 200 HZ to 80 HZ. This kept the bass localised to the front image.

own stereo reverb sub-master, reducing the stereo mix to two faders, dry band and reverb — the *sound-source* and *sound-space.* These were then routed to a full-mix sub-master and lastly to the full-mix-print track. For construction of the 5.1 *sound-space,* three Waves 5.1 R-360 reverbs were used, each with different reflection and reverb traits, and an additional EMT 250, with a 60-millisecond pre-delay and 2.2 seconds of reverb length was blended into the LS/RS. This accomplished the perception of reverb travel from front to rear in the *sound-space*. It formed the rear spatial-counterpoint to the original stereo *sound-space* which remained in the L/R front and added some commonality between the *sound-field* and *sound-scape* versions. Together, all these reverbs created a thick and complex *sound-space.* Global reverb EQ involved rolling off high frequencies and emphasizing the 200 to 800 Hz range of the audio spectrum by +2.5 db. A very small amount of the outside piano microphones and the front bass hyper-cardioid mic were blended into the center channel for *sound-field* support, creating a pleasing and somewhat immersive L/C/R sound. The inclusion of a discrete center channel served to eliminate the phantom center of a L/R *sound-field,* providing more accurate positioning in the horizontal panning (image 3).

Mix philosophy

To create a virtual *sound-field* or *sound-scape*, a mix engineer must combine analytical listening, personal sonic aesthetics and audio engineering skills. Composer R. Murray Shafer's thoughts about *soundscape* propounded in his book *The Tuning of The World*. (Schafer: 1977) are both illuminating and revelatory:

> "What the soundscape analyst must do first is to discover the significant features of the soundscape, those sounds which are important either because of their individuality, their numerousness or their domination."
>
> "Outdoor sounds are different than indoor sounds. Even the same sound is modified when it changes spaces. The human voice is always raised outdoors."
>
> "When there is no sound, hearing is most alert.

These thoughts can help mixers to understand acoustic environment by providing insight into the extended psychological and cultural meaning of their decisions. For example, a jazz bass drum does not sound like an electronic dance music bass drum. Surround-sound introduces aesthetic and technical considerations that are more complex than stereo. I believe that awareness of Shafer's soundscape philosophy can assist a mixer to evolve a *sound-field* into a convincing *sound-scape* which would be heard as a convincing "virtual acoustic environment."

Mix techniques: Piano

The piano width was recorded intimately, using an inside-placed Calrec-Soundfield (C-SF) microphone. B-channel signal (WXYZ) was not used but the mic was placed about nine inches above the strings and directly over the Yamaha C-7 logo on the soundboard. The cardioid pickup pattern was set to about 1:30 on the rotary dial of the C-SF SPS422B pre-amp and the width was set to ultra-wide. The two outer Neumann U87s formed more of a mono image. It's essential for the orientation of the L/R piano image to be heard as if one were seated at the instrument: low notes on the left, high notes on the right. The U87s were predominant in the piano blend while the C-SF image was panned slightly inward, reserving the outer edges of the L/R *sound-field* for the virtual *sound-space*.

Bass

Creating a stereo version first on either a 2 channel or a 2.1 bass managed speaker system provides an excellent foundation for fold-out to 5.1. Experience has shown me that fine-alignment of an LFE subwoofer is not an exact science, but if a stereo mix plays with confidence in bass-managed 2.1, then sub-woofer alignment in 5.1 should be close. Unique to the *sound-scape,* some D.I. was routed directly to the LFE to tighten-up the bass sound. The C-391 provided the focus and the AEA ribbon added spatial texture.

Reference

For a comparative mixing reference, I chose two songs, *Daydream* and *The Oracle*, from Kenny Barron and Dave Holland's duo recording *The Art of Conversation* because I liked their ensemble sound. Izotope's *Insight* plug-in provided analysis of their levels, tone balance and *sound-field*.[15] Since these examples were mastered at approximately -16 LUFS, the average stereo mix level of *Look Ahead* was reduced by about four decibels (db) to provide extra dynamic range. This four db was carefully regained during mastering.

Automation

Instead of fader automation, clip-gain was the primary edit tool for blending. This technique sent consistent levels to compressor thresholds, keeping the mix dynamically open and even.

[15] Barron, Kenny, Holland, Dave, *The Art of Conversation*, Izotope analysis: Daydream, -19.3 LUFS Left Peak -1.0, Right Peak -0.1 | The Oracle -15.9 LUFS—Left Peak-0.3. Right Peak -0.2

Reverb sends were routed as pre-fader, so the ratio of dry-to-wet could be easily auditioned and automated to change the balance between wet and dry. The sound of stereo reverb panned across four or five speakers is not pleasing because it is unidimensional. The Waves R-360 5.1 reverb solved this by joining the front-stereo and rear-stereo reverbs together with diverse multichannel reverb signals, eliminating the "reverb dead-zone" on each side between the left/right front and LS/RS rear speakers. This presented a multidimensional *sound-space.*

Roll off

The stereo version of *Look Ahead* was mixed first and then the 5.1 version was created, maintaining a universal balance of *sound-source* and *sound-space*. The sound-space employed an LFE (low frequency effects) LPF (low pass filter) roll-off of 24 db per octave started at about 80 HZ and the LS/RS channels began their HPF (high pass filter) roll off at about 200 Hz, ending at 80 HZ. This crossover was done with the Waves 360 manager, and helped to keep the staging focused on the front L/C/R audio image.

Unique mix opportunities in 5.1

The 5.1 *sound-scape* presented several other unique opportunities such as custom re-balancing of the *sound-space* with the *sound-source* at opportunistic moments. For example, on *Now's the Time,* at approximately forty-one seconds, a hand-on-piano string muting technique with careful pedaling was used to produce a percussive and sustaining string effect. This pleasing sound benefited from an increased level of *sound-space* that created an immersive, mild, slow chorusing effect. Example two *(see appendix)* focuses on this section and peels back the layers of *sound-source*, *sound-space*, *sound-field* and *sound-scape* in a side-by-side comparison.

The 5.1 *sound-space* was exaggerated on the endings of *Praise* and *A Nightingale Sang in Berkeley Square* — the chamber level rises by about twenty percent, creating a haunting goodbye as the dry-band fades into the distant and immersive reverb chamber.

Arco bass solo moves from a public to an intimate space

Also unique to the *sound-scape*, the introduction on *Porters Hymn* includes an arco bass solo, set far back on the left side of the virtual stage. When the pizzicato melody enters, the perspective returns to the front center, illustrating Moore's "soundbox model" (Moore: 1992) by first presenting the bass in a distant public space and then quickly moving it to an intimate position.

Mastering

Mastering brought the LUFS level up to about -16.0 with peaks limited to -0.7. U.A. mastering plugins such as the Manley Massive Passive, Ampex ATR-102, Millenia NSEQ-2, and the Precision Limiter improved the tonal focus, adding saturation and peak limiting while extending the high and low end spectral range.

Reflective thoughts lead to ITPRA theory

For fifteen years, I have folded-down 5.1 to stereo in broadcast post-production, but this is the first time that I folded-out. I now see that it allows unique decisions to be made that can enliven emotional engagement in the *sound-scape*, while maintaining the ensemble from the pre-existing *sound-field*. However, I think that it's still important for surround-sound music creators to observe how film/broadcast and media-arts will stage their sound in future multichannel formats, notably how NHK's upcoming convention—8K Super Hi-Vision with 22.2 audio—will provide home-theatre audiences with the ability to select language, adjust dialogue volume, and hear audio-objects in immersive 3D.[16] New technology brings challenge and opportunity by affording multichannel music creators more options for playback, at the same time making it important to respect the disciplined efforts of those who follow the pedagogy of composition, orchestration and arranging to pursue cohesive music ensemble. Conversely, 22.2 multichannel home theatre audio may also inspire new possibilities for compositions and audio productions that aim for hyper-realistic *sound-scapes*.[17]

Starting with *Fantasia* in 1940, film audio history shows an evolution of technology and sound that tried to further stimulate human emotional expectation and experience with immersive *sound-scape*. Walt Disney wanted to choreograph the musical-sound of a bumblebee twirling around his audience to heighten their emotional experience. Maestro Leopold Stokowski suggested that Disney engineers speak with Bell Labs about their research in multi-channel audio, leading to *Fantasound*. Soon after that, in the '50s, emotional experience induced by music was explored in Leonard Meyer's seminal book, *Emotion and Meaning in Music*, asserting that music communicates emotion by "choreographing expectation" (Meyer: 1956).

In 2006, David Huron's book, *Sweet Anticipation*, identified the *ITPRA theory*, a psychological theory of expectation (Huron: 2006). ITPRA aims to better understand musical tension and release by identifying five internal

[16] *8K, Super Hi Vision,* involves a 22.2ch. sound microphone, portable recording and editing device, 3D audio mixing system, mixer for live feeds, 3D reverberator, 22.2ch. headphones, processor etc.

[17] Lopez, Francisco, Electro Acoustic artist whose music style is based on immersive soundscape, inspired by R. Murray Schafer's book, *The Tuning of The World,* https://open.spotify.com/artist/2NIo9CSlxx2pArdKee4hes

emotion-response systems: 1) *Imagination*, 2) *Tension*, 3) *Prediction*, 4) *Reaction* and 5) *Appraisal*. ITPRA can be grouped into two categories of *pre-outcome tension* and *post-outcome release*. The first three *pre-outcome* states (ITP) describe how anticipation builds toward a stressful level of tension, and the last two *post-outcome* states (RA) describe the release of that tension. Decisions about meter, syncopation, tonality, spatiality and cadence leverage the ITPRA theory of expectation to elicit an arc of emotional responses from the audience. Expanding upon Meyers observations, I now view creators of *sound-field* and *sound-scape* as "choreographers of emotion and expectation." Their individual sense of ITPRA guides the organization of sound, which can be expressed anywhere between the virtual contexts of a small room, to a natural outdoor setting or, even further outward, to an interstellar space battle portrayed with cinematic exuberance.[18] I suspect that the distinctive traits of immersive surround-sound could even help canny record producers of concert music to find their particular listening audience, without allowing the technology to deconstruct the ensemble or change its focus on the performance and composition.

Performance touches the sound-space

Look Ahead arrives at its performative destinations via the virtual acoustic environment of an indoor concert chamber with staging that supports intimate music. Schafer's statements: "What the soundscape analyst must do first is to discover the significant features of the soundscape" and "When there is no sound, hearing is most alert" proved to be foundational in creation of that virtual acoustic environment. The dynamic of this recording resides predominantly between *pianissimo* and *mezzo-forte* and rarely goes beyond *forte,* this trait invites the condition of alert hearing that Schafer speaks of. A live-performance dynamic is foundational to the mix because it excites the *sound-space*. A multi-speaker listening environment can support quiet dynamics very well due to the diffused or "spread-out" channel energy that naturally occurs from a multi-speaker installation. In *Look Ahead*, surround-sound never draws attention to itself by panning an instrument to an unexpected place. This encourages the listener to suspend their believe in the virtual *sound-scape.* Trusting that the listener would become solidly immersed in this virtual sound, I increased *sound-space* level by about 2-db for *Praise* and *Budapest*, because the slow tempo—duration between the beats—allowed the chamber to ring clearly without "smearing." These specific adaptations also illustrate Byrne's theory of creation in reverse as they

[18] Bhatia, Amin, *The Interstellar Suite*, Produced by Frank Marrone, Triplet Records, 2015, TR-10018

did not occur to me while working in the *sound-field*, but the *sound-scape* called out for them.

Supporting examples and data analysis

Example #1: A side-by-side comparison of a reverse engineered fold-down from the 5.1 master and the prepared stereo 2.0 master. It illustrates how the static convergence of LS/RS results in an inconsistent *sound-space* that portrays undesirable and varying levels of intimacy. The bass solo of *My Favourite Things* was selected to show these undesirable results. (see appendix)

Example #2: Reveals the audio-channel layers from stereo *sound-source* and *sound-space* to *sound-field* and *sound-scape* in a side-by-side playback of the piano hand-mute section of *Now's the Time.* (see appendix)

Data analysis

While both prepared masters of *My Favourite Things* sound entirely different, they share common cohesive ensemble and loudness data. Image 4 provides a comparison of the LUFS and peaks between the fold-down to 2.0 and the prepared 2.0. The numbers are remarkably close, and this is representative of the entire album. However, the masters sound very different, proving that one must always listen and not just look at meters.

Image 4: My Favourite Things, side-by-side analysis of levels. Left: Prepared stereo 2.0 & Right: 5.1-Fold-down to stereo 2.0.

The prepared stereo analysis is displayed on the left instance of Izotope *Insight* and the fold-down and its down-mixer settings are on the right *(image 4)*.

Conclusions from research and practice

The practice and its four selections of research: 1) Surround-sound history; 2) David Byrne's theory of "creation in reverse"; 3) R. Murray's Schafer's theories of soundscape analysis; and 4) David Huron's general theory of human expectation—ITPRA, distill into seven foundational guiding insights that may assist the creator when adapting *sound-field* to a *sound-scape.*

1. Since 1881, multichannel audio has been heightening audience experience and it continues to do so, suggesting that immersive audio should not be overlooked by music creators. It's possible that music *sound-scape* could become a more significant music distribution format because it seeks an expert listener who desires a premium product.
2. Creative decisions about how to utilize multichannel contexts are informed by the available technology which is evolving constantly.
3. These pre-existing contexts determine what the reception of immersive experience can be. They range from a re-creation of a natural acoustic environment to a hyper-real *sound-scape.*
4. Acoustic environments—soundscapes—influence and inform the creation of virtual *sound-fields* and *sound-scapes.*
5. Composers and creators of immersive sound can be considered *"choreographers of emotion and expectation."*
6. Even though data analysis of loudness and peaks can show very little variance between a fold-down to 2.0 stereo and the prepared 2.0 stereo version, the virtual acoustic environments of *sound-field* and *sound-scape* can sound very different, resulting unique listening experiences and emotional outcomes.
7. Adapting a stereo *sound-field* 2.0 mix from its mix template into a *sound-scape* is a preferred method for two reasons: 1) It results in two individually prepared masters that share cohesive ensemble in their unique virtual *sound-spaces*; and 2) a fold-down can result in unpredictable changes to depth perspective when listening for the sounds of intimacy, personal, quasi-public, and public spatiality in the down-mixed *sound-field.*

Closing thoughts

My first surround-sound playback experience taught me not to confuse the listener by pulling an ensemble apart. A creator needs to understand the aesthetic essence of their music and then integrate it with the pre-existing technological formats accordingly. Once the playback formats are chosen, accepted engineering practice indicates that the ideal way to organize the sound is with a 'film-style' mixing template because it provides accessibility to individual instrument tracks that can be routed (bussed) to stereo and multichannel sub-masters.

The essential take away is that a fold-out from a *sound-field* can provide a cohesive foundation that confidently leads to a *sound-scape*. The result is two diverse masters that share a common and familiar ensemble sound. For the listener of both, these commonalities and differences between *sound-field* and *sound-scape* may even help to keep the music fresh regardless of how many times it's listened to.

I hope that this case study inspires practitioners of the recording arts to experiment with immersive music production while maintaining ensemble sound. Hopefully, students and professionals will consider these insights when they adapt their own music from stereo *sound-field* to immersive *sound-scape*.

Acknowledgments

Thanks to Canada Council for the Arts, Robi Botos, Amin Bhatia, Frank Morrone, David Greene, Jeff Wolpert, Tom Reynolds, Harry Lewis, Michael Francis Nunan, Hayward Parrott, Chad Irschick, Doug Wilde, Matt Zimbell, Peter Cook, Steve Bellamy, Bruce Cassidy, Michael Colgrass, Barry Elmes, Al Henderson, Rob van der Bliek, Jay Rhan, Michael Coghlan, Dr. Norman Ludwin, Keith Loach, George Novotny, Adrian Novotny and Dave Dysart for the many thoughtful discussions regarding practice, study and analysis of music composition, recording, engineering, mixing, film, broadcast and audio post. Special thanks to Universal Audio, HHB Canada, Yamaha, Humber College and York University for their continued support. I acknowledge the support of the Canada Council for the Arts, which last year invested $153 million to bring the arts to Canadians throughout the country.

References

Bibliography

Adorno, T. (1976) Introduction to the sociology of music, translated by E.B. Ashton, Seabury Press, New York

Blumlein, A.D. (1931) "British Patent Specification 394, 325," reprinted in the Journal of the AES, (1958) 6, p. 91 and in the JAES anthology Stereophonic Techniques.

Byrne, D. (2012) How Music Works, McSweeney's Books, Pg.18

Dientsfry, E. (2016) The Myth of The Speakers: A Critical Re-examination of Dolby History, Film history, Indiana University Press, 28, 1, pp. 167-193, www.jstor.org/stable/10.2979/filmhistory.28.1.06

Ellis, C. Adams, T. Bochner, A. (2011) 'Autoethnography: an overview', Historical Social Research/Historische Sozialforschung, 36, 4, pp. 273-290
Gerzon, M.A. (1992) Applications of Blumlein Shuffling to Stereo Microphone Techniques, Journal of the AES, #3448
Huron, D. (2006) Sweet Anticipation, Music and the Psychology of Expectation. The MIT Press, Cambridge, Massachusetts, London England, pp. 33-34, iBooks
Julstrom, S. (1987) A High-Performance Surround-Sound Process for Home Video, Sure Brothers Inc., Evanston, Il 60202, USA, Journal of the AES, #5192, 35, 7/8-July/August
Klapholz, J. (1991) Fantasia—Innovations in Sound, Journal of the AES, #5994, 39
Meyer, L,B, (1956) Emotion and Meaning in Music, University of Chicago Press
Moore, A. (1992) Rock: The Primary Text, Developing a musicology of rock, Open University Press, Buckingham, England
Moore, A. (2012) Song means: Analysing and interpreting recorded popular song, Ashgate Publishing, Kindle edition, (2016) by Routledege, Milton Park, Abingdon, Oxon & NewYork
Moore, A. (2018) Rock: The Primary Text, Developing a musicology of rock, Revised Kindle edition, Routledge, Milton Park, Abingdon, Oxon & NewYork
Moylan, W. (2002) The Art of Recording, understanding and crafting the mix (Boston MA: Focal Press
Pickering, N. Baender, E. (1953) Two Ears in Three Dimensions, Journal of the AES, #94, 1,3
Schafer, Murray, R: (1977) The Tuning of the World, - Alfred A. Knopf Publishing, reprinted (2012) in The Soundscape, Johnathan Sterne, Taylor and Francis
Sigimoto, Takehero, Nkayama, Yasushige, Komori, Tomoyasu, Chinen, Toru, Hatanaka, Mitsuyuki: (2017) Dialogue Channel Control for 22.2 Multichannel Sound Broadcasting: Broadcast Chain Scheme and Subjective Evaluation of Effectiveness, Journal of the AES D-00132_HR_22.2 sound, 65, 6, Twitter-@8kSHV
Torick, E. (1998) Highlights in the History of Multichannel Sound, Journal of the AES, 46, 1/2 January/February, pp. 27-31
Unknown author, (1881) Scientific American, Hertz, B.F. (1981): Engineering - 100 Years with Stereo: The Beginning, reprinted in the Journal of the AES, PDF #18964, 29, 5, pp. 368-372

Discography

Barron, Kenny, Holland, Dave, The Art of Conversation, Impulse Records, 2014
Bhatia, Amin, The Interstellar Suite, Triplet Records, 2015
Lopez, Francisco, https://open.spotify.com/artist/2NIo9CSlxx2pArdKee4hes
Novotny, Paul, Botos, Robi, Look Ahead, Triplet Records, 10019, 2016

Websites

https://www.dolby.com/us/en/guide/speaker-setup/index.html-Accessed 11-02-17
https://www.dolby.com/us/en/guide/test-tones.html (Windows and Mac)-Accessed 11-02-17
http://techland.time.com/2013/06/06/dolby-atmos/-Accessed 11 13 17
https://media.netflix.com/en/company-blog/dolby-atmos-coming-to-netflix-Accessed 09-05-17
https://prezi.com/qgew4ppz7kng/history-of-surround-sound-Accessed 04-28-17
https://en.wikipedia.org/wiki/Surround_sound-Accessed 11-10-17
http://www.nhk.or.jp/8k/index_e.html-Accessed 11-8-17
https://en.wikipedia.org/wiki/DolbyAtmos-Accessed 11-13-17
http://investor.dolby.com/releasedetail.cfm?ReleaseID=799045-Accessed 11-13-17
https://www.dolby.com/us/en/technologies/music/dolby-atmos.html-Accessed 11-24-17
http://www.thebsa.org.uk/wp-content/uploads/2014/04/soundfieldguidelinesfeb2008.pdf-Accessed 03-18-18
http://www.holophone.com/default.asp-Acessed 03-19-18
http://www.telephonecollecting.org/Bobs%20phones/Pages/Essays/Ader/Ader.htm 04-25-18

Terminology

Sound-field: A stereophonic recording played back presents an *acoustical environment* with the sound coming from front L/R speakers.

Sound-scape: An immersive multichannel recording played back presents an *acoustical environment* with the sound coming from front L/C/R speakers and rear side, LS/ RS with an LFE (low frequency effects) channel.

Soundscape: R. Murray Schafer's term for *acoustical environment.* David Byrne uses the term *acoustic context,* and I interpret that he means *acoustical environment.*

Sound-space: The reverb of a sound-field or sound-scape. A term presented by Emil Kraugerud in his JARP paper, *Meanings of Spatial Formation in Recorded Sound,* ISSN:1754-9892 | March 2017.

Sound-source: The dry audio source containing no virtual reflection, echo or reverb.

LUFS: Loudness Units Full Scale.

Appendix

Analysis example #1: Side-by-side fold-down analysis–5.1 to 2.0 compared to the stereo master, illustrating why fold-out is necessary, Bass solo from My Favourite Things, https://www.youtube.com/watch?v=fGZu2w1gZIQ

Analysis example #2: Unpacking the Sound-scape_2.0 Sound-field to 5.1 Sound-scape, Now's the Time, https://www.youtube.com/watch?v=93tagVfUviQ&t=15s

Look Ahead in 24/96 lossless Flac, 5.1 sound-scape and stereo 2.0 sound-field high-resaudio.com/en/album/view/dpdwkx/paul-novotny-robi-botos-look-ahead

Look Ahead in 2.0 stereo, 24/44.1 Mastered for iTunes, MFIThttps://itunes.apple.com/ca/album/look-ahead/1113854143